


## RESEARCH ARTICLE

# Geostatistical models using remotely-sensed data predict savanna tsetse decline across the interface between protected and unprotected areas in Serengeti, Tanzania

Jennifer S. Lord<sup>1</sup>  | Stephen J. Torr<sup>1</sup> | Harriet K. Auty<sup>2</sup> | Paddy M. Brock<sup>3</sup> |  
 Mechtilda Byamungu<sup>4</sup> | John W. Hargrove<sup>5</sup> | Liam J. Morrison<sup>6</sup> | Furaha Mramba<sup>4</sup> |  
 Glyn A. Vale<sup>5,7</sup> | Michelle C. Stanton<sup>1</sup>

<sup>1</sup>Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK; <sup>2</sup>Epidemiology Research Unit, SRUC, Inverness, UK; <sup>3</sup>Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK; <sup>4</sup>Vector and Vector-Borne Diseases Research Institute, Tanga, Tanzania; <sup>5</sup>SACEMA, University of Stellenbosch, Stellenbosch, South Africa; <sup>6</sup>Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK and <sup>7</sup>Natural Resources Institute, University of Greenwich, Chatham, UK

## Correspondence

Jennifer S. Lord

Email: jennifer.suzanne.lord@gmail.com

## Funding information

Zoonosis and Emerging and Livestock Systems (ZELS) programme, Grant/Award Number: BB/L019035/1; UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), Grant/Award Number: 221948, ICONZ; Biotechnology and Biological Sciences Research Council; Department for International Development; The Economic and Social Science Research Council; The Natural Environment Research Council and the Defence, Science and Technology Laboratory; Canadian International Development Research Centre (IDRC); European Union's Seventh Framework Program, Grant/Award Number: FP7/2007-2013; ICONZ (Integrated Control of Neglected Zoonoses)

Handling Editor: Izak Smit

## Abstract

1. Monitoring abundance is essential for vector management, but it is often only possible in a fraction of managed areas. For vector control programmes, sampling to estimate abundance is usually carried out at a local-scale (10s km<sup>2</sup>), while interventions often extend across 100s km<sup>2</sup>. Geostatistical models have been used to interpolate between points where data are available, but this still requires costly sampling across the entire area of interest. Instead, we used geostatistical models to predict local-scale spatial variation in the abundance of tsetse—vectors of human and animal African trypanosomes—beyond the spatial extent of data to which models were fitted, in Serengeti, Tanzania.
2. We sampled *Glossina swynnertoni* and *Glossina pallidipes* >10 km inside the Serengeti National Park (SNP) and along four transects extending into areas where humans and livestock live. We fitted geostatistical models to data >10 km inside the SNP to produce maps of abundance for the entire region, including unprotected areas.
3. Inside the SNP, the mean number of *G. pallidipes* caught per trap per day in dense woodland was 166 (± 24 SE), compared to 3 (±1) in grassland. *Glossina swynnertoni* was more homogenous with respective means of 15 (±3) and 15 (±8). In general, models predicted a decline in abundance from protected to unprotected areas, related to anthropogenic changes to vegetation, which was confirmed during field survey.
4. *Synthesis and applications.* Our approach allows vector control managers to identify sites predicted to have relatively high tsetse abundance, and therefore to design and implement improved surveillance strategies. In East and Southern Africa, trypanosomiasis is associated with wilderness areas. Our study identified pockets of vegetation which could sustain tsetse populations in farming areas outside the

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Journal of Applied Ecology* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

Serengeti National Park. Our method will assist countries in identifying, monitoring and, if necessary, controlling tsetse in trypanosomiasis foci. This has specific application to tsetse, but the approach could also be developed for vectors of other pathogens.

#### KEYWORDS

disease, geostatistical models, *Glossina*, pathogens, protected areas, remote-sensing, surveillance, trypanosomiasis, tsetse, vector control

## 1 | INTRODUCTION

In sub-Saharan Africa, the edges of protected areas, including national parks and game reserves, have experienced human population growth twice that of other rural areas (Wittemyer et al., 2008). These regions present a complex transmission context for zoonotic diseases, including Rhodesian human African trypanosomiasis (r-HAT; Auty, Morrison, Torr, & Lord, 2016; Hassell, Begon, Ward, & Fèvre, 2016). In these areas, *Trypanosoma brucei rhodesiense*—the causative agent of r-HAT—circulates in wildlife and livestock, alongside the trypanosome species which cause animal African trypanosomiasis (AAT). A better understanding of trypanosome transmission dynamics and approaches to improve control are current priorities for these regions (Auty et al., 2016; Diall et al., 2017).

Tsetse control is the only method available for reducing r-HAT within and around the edges of protected areas. Savanna tsetse—the Morsitans group of *Glossina*—are the primary vectors of the trypanosomes which cause r-HAT and AAT inside and adjacent to protected areas in east and southern Africa (Gondwe et al., 2009; Mweempwa et al., 2015). For livestock-owning communities, use of insecticide-treated cattle is the most cost-effective method of tsetse control but it requires sufficient densities of cattle (Shaw et al., 2015). Where cattle are sparse, insecticide-impregnated targets baited with attractants can be used. Application of these vector control methods over large and remote protected areas is not feasible, but control can be targeted in interface areas to reduce human and livestock exposure. In order to concentrate resources for optimal cost-effective control requires information on the distribution and abundance of tsetse at a local level—10s km<sup>2</sup>, but district/country-level maps are often only available (Albert, Wardrop, Atkinson, Torr, & Welburn, 2015; Dicko et al., 2014; Hendrickx et al., 1999; Wint & Rogers, 2000). Although atlases of HAT and AAT exist (Cecchi et al., 2014; Simarro et al., 2010), for Tanzania, local-scale and contemporary knowledge of tsetse distribution and abundance is lacking.

Extensive grid-based sampling of tsetse using traps to estimate abundance coupled with vegetation mapping using remotely-sensed data has previously been applied to inform tsetse control (Albert et al., 2015; Dicko et al., 2014; Hendrickx et al., 1999). However, such sampling is logistically intensive and expensive. This is particularly so within or at the edges of protected areas which are often remote and difficult to access. The ability to predict where vegetation is suitable for tsetse, using remotely-sensed data, would be valuable for

informing surveillance and control (Kalluri, Gilruth, Rogers, & Szczur, 2007).

Remotely-sensed data can be included in geostatistical models to identify areas where vegetation may be suitable for tsetse (Albert et al., 2015; Bouyer et al., 2010; Dicko et al., 2014; Ducheyne et al., 2009; Kitron et al., 1996; Mweempwa et al., 2015). However, few studies have linked tsetse abundance, habitat and remotely-sensed variables on a local scale (Kitron et al., 1996). Moreover, local-scale statistical models for tsetse have not been tested for their ability to predict abundance in regions other than those for which the original model was produced (Albert et al., 2015; Bouyer et al., 2010; Dicko et al., 2014; Ducheyne et al., 2009; Kitron et al., 1996; Mweempwa et al., 2015).

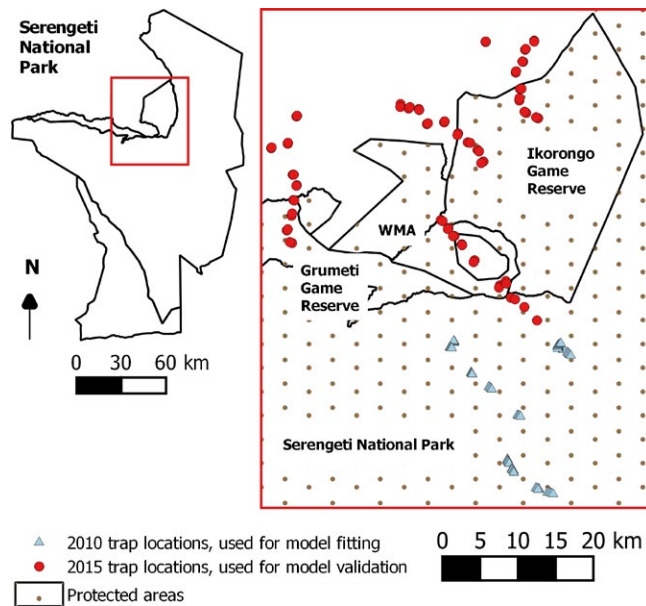
In regions >10 km inside protected areas, the drivers of tsetse population dynamics are limited to be natural variation in vegetation and wildlife densities associated with vegetation (Allsopp, Baldry, & Rodrigues, 1972). Rather than interpolation (Albert et al., 2015; Dicko et al., 2014; Hendrickx et al., 1999), comparing extrapolated predictions from models fitted to data inside protected areas with data from across the interface would be a way of testing the robustness of relationships between remotely-sensed data as indicators of habitat and tsetse abundance. It may also allow insight into the drivers of tsetse distribution at the interface between protected and unprotected areas (Miller, Turner, Smithwick, Dent, & Stanley, 2004).

We quantified mean daily numbers of tsetse caught in traps as a function of vegetation type, for areas >10 km inside the Serengeti National Park (SNP), Tanzania. We then used geostatistical models based on remotely-sensed data fitted to these tsetse catches to predict how abundance varies spatially in regions across the interface between protected and unprotected areas. We tested model predictions with new field data from the interface.

## 2 | MATERIALS AND METHODS

### 2.1 | Study area

The protected areas in the study, comprising approximately 3,000 km<sup>2</sup> (Figure 1) support a gradient of habitats from dense woodland to grassland (Reed, Anderson, Dempewolf, Metzger, & Serneels, 2009). To the northwest of the SNP, communities practice livestock keeping and mixed crop-livestock farming. Some unprotected areas here however still support natural vegetation (Estes, Kuemmerle, Kushnir,



**FIGURE 1** Study area and Nzi trap locations. Triangles: 2010 Nzi trap sampling sites >10 km inside the Serengeti National Park. Circles: 2015 Nzi trap sampling sites across the interface between protected and unprotected areas. WMA, wildlife management area

Radeloff, & Shugart, 2012). The northern region of Tanzania, including the SNP, has been known as an r-HAT focus since the 1920s (Davey, 1924; Swynnerton, 1923). More recently, human infective *T. brucei rhodesiense* has been identified by PCR in cattle and wildlife in the study area (Auty, 2009; Kaare et al., 2007). Although r-HAT cases in residents are likely underreported (Odiit et al., 2005), 30 cases in travellers to the region were reported between 1990 and 2007 (Auty, 2009). The SNP supports three species of tsetse: *Glossina swynnertoni*, *Glossina pallidipes* and *Glossina brevipalpis*—although the last species is present in smaller numbers than Morsitans group species (Auty et al., 2012).

## 2.2 | Quantifying tsetse abundance

Surveys were carried out in February 2010 >10 km inside the SNP and in February 2015 across the interface between protected and unprotected areas (Figure 1). Sampling was carried out in February, prior to the rains to minimise water damage to traps and because tsetse abundance in Tanzania has previously been found to be highest between February and June (Glasgow & Welch, 1962). Tsetse were sampled using Nzi traps (Mihok, 2002) baited with a blend of acetone (100 mg/hr), 1-octen-3-ol (0.5 mg/hr), 4-methylphenol (1 mg/hr) and 3-n-propylphenol (0.1 mg/hr; Torr, Hall, Phelps, & Vale, 1997). Trap locations were recorded using a GPS. Catches from traps were collected daily for three consecutive days and the numbers of tsetse of each sex and species were recorded. We use the term “abundance” to describe mean numbers of tsetse caught per trap per day.

In February 2010, traps were deployed in each of four vegetation types, categorised, using an existing vegetation map (Hopcraft, 2008)

and ground-truthing, as grassland (<2% tree cover), savanna (2%–20% tree cover), open woodland (20%–50% tree cover) and dense woodland (50%–100% tree cover). Traps were deployed along three transects (Figure 1) extending from c. 10 to c. 30 km inside the SNP, based on accessibility from roads. Eighteen traps were placed in dense woodland, 14 in grassland, 14 in open woodland and 14 in savanna. Traps in each vegetation type were at least 50 m apart in dense/open woodland or 100 m apart in grassland/savanna, because traps are more visible to tsetse in open areas.

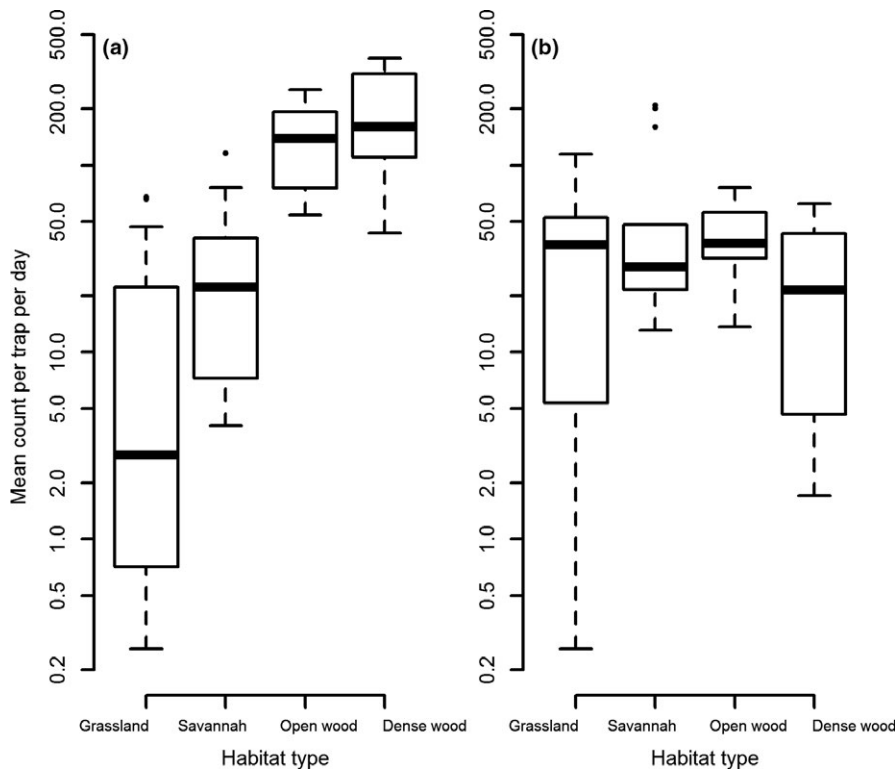
In February 2015, four transects of Nzi traps were set from 5 km inside, up to 10 km outside the protected area boundary (Figure 1). Transects were selected based on their accessibility from roads. Along each transect, we set pairs of odour-baited Nzi traps at approximately 1.5-km intervals irrespective of vegetation type. The Euclidean distance of traps from the protected area boundary was estimated.

We assume that our trapping did not affect the tsetse population. An odour-baited insecticide-treated target, which will catch similar numbers of tsetse as an odour-baited trap, kills c. 1% of tsetse per day within a km<sup>2</sup> (Vale, Hargrove, Cockbill, & Phelps, 1986). To have an impact on a tsetse population requires >3% of the population within a km<sup>2</sup> to be captured or killed per day and this has to be applied for months to impact the population (Hargrove, 1988).

## 2.3 | Remotely-sensed variables

Numbers of *G. pallidipes* caught in traps have previously correlated inversely with Landsat shortwave infrared values—an indicator of moisture (Barsi, Lee, Kvaran, Markham, & Pedelty, 2014; Kitron et al., 1996). Normalized Difference Vegetation Index (NDVI) values above 0.39 have previously been used as an indicator of vegetation suitable for tsetse (Lin, DeVisser, & Messina, 2015; Moore & Messina, 2010), based on the observation that mortality rates decrease as NDVI increases (Rogers & Randolph, 1991). The NDVI is a measure of the density of plant matter—using the near-infrared and visible red wavelengths. We also included elevation and land surface temperature (LST), given the importance of temperature to tsetse mortality rates (Hargrove, 2004). For estimates of LST, the Landsat 8 Thermal Infrared Band 10 image was converted to at-satellite brightness temperature. We then used NDVI to estimate emissivity and calculate LST as previously described (Sobrino, Jiménez-Muñoz, & Paolini, 2004). For elevation, we used the ASTER Global Digital Elevation Model (GDEM), which is a product of NASA and METI and has a resolution of 30 × 30 m.

Three Landsat 8 images (30 × 30 m resolution) with less than 10% cloud cover—Path/Row 169/061, 169/062 and 170/061—and the GDEM were acquired from Earth Explorer (<https://earthexplorer.usgs.gov/>) from 13 February 2015 and 20 February 2015. The 2015 images had lower cloud cover than that available from 30 January 2010, and both sets of images appeared sufficiently similar to use the 2015 set for both time points. Histogram matching, conversion to near-surface reflectance, calculation of NDVI and application of a cloud mask was done in R (R Core Team, 2014).



**FIGURE 2** Number of tsetse caught per trap per day by habitat type inside the Serengeti National Park, February 2010. Y-axes on a log scale. (a) *Glossina pallidipes* was significantly lower in grassland and savanna than in open and dense woodland and was lower in grassland compared with savanna (ANOVA  $F_{38.46}$ ,  $p < .001$ , Tukey HSD  $p < .001$  for each pairing with either grassland or savanna,  $p = .02$  for grassland-savanna); (b) *Glossina swynnertoni*—no significant difference between habitat types ( $p > .05$ )

## 2.4 | Data analysis

Only 56 *G. brevipalpis* were caught throughout our studies and numbers were too low for analysis. *Glossina pallidipes* and *G. swynnertoni* count data were overdispersed. To compensate for this, daily catches ( $y$ ) from 3 days for each trap were transformed to  $\log_{10}(y + 1)$  before calculation of mean numbers per trap per day and SE.

For each species, we first explored the variation in numbers per trap per day, >10 km inside the SNP during February 2010, between vegetation types. To determine the significance of the difference in log-transformed abundance observed between vegetation types for each species, we used ANOVA followed by the Tukey Honest Significant Difference test (Tukey HSD). We then fitted a Bayesian linear geostatistical model (Brown, 2015; Diggle, Tawn, & Moyeed, 1998) only to the February 2010 data from >10 km inside the SNP. The function `glgm` within `geostatsp` (Brown, 2015) was used as it provided a user-friendly method for fitting and comparing models within a Bayesian framework. The `glgm` function implements integrated nested Laplace approximation (Rue, Martino, & Nicolas, 2009) for fitting and prediction and outputs cross-validation measures for model comparison.

Allowing for savanna tsetse daily dispersal rates (Hargrove, 1981), a buffer of 500 m around each trap was used to calculate the average LST, Landsat Band 7, NDVI and elevation. We used a systematic model fitting approach focused on the predictive ability of the model, rather than identifying a causal relationship between the environmental variables and the trap data. Separate linear regression models were first fitted using each of the four individual

environmental variables. Model fits, initially without a spatially explicit error term, to the *G. pallidipes* and *G. swynnertoni* data, were compared using the negative of the sum of the log conditional predictive ordinates (log-CPO score; Held, Schrödle, & Rue, 2010). Conditional predictive ordinates indicate the ability of a model to predict the value of a data point omitted from the fitting step. Starting with the univariate model with the lowest log-CPO score, each of the remaining variables was added to the model one at a time. The log-CPO was recalculated with each additional variable, and the combination with the lowest log-CPO was determined to be the most parsimonious.

Rasters of environmental variables remaining in the most parsimonious model were resampled to a resolution of 500 × 500 m, considering tsetse daily dispersal rates and to reduce computing time, covering an area including both 2010 transects inside the SNP and 2015 transects across the interface. We then compared the fit, to the 2010 data, of a geostatistical model including a spatially structured error term to a model including an unstructured error term by comparing log-CPO scores. Normally distributed errors were assumed, and prior 95% intervals for the practical range were 500 and 5,000 m.

To test the ability of our models to predict tsetse abundance, predictive maps were simultaneously produced.  $\log_{10}(y + 1)$  transformed geostatistical model predictions for new locations sampled during 2015 across the interface were then compared with  $\log_{10}(y + 1)$  transformed observed values and residuals between predicted and observed counts calculated.

All data and R scripts required to produce the figures are available online (Lord et al., 2018).

### 3 | RESULTS

#### 3.1 | Tsetse abundance >10 km inside the SNP

During 2010 surveys >10 km inside the SNP, mean numbers of *G. pallidipes* per trap per day were 123 ( $\pm 24$ ) in open woodland, and 166

**TABLE 1** Negative of the sum of the log conditional predictive ordinates (log-CPO score) for linear models using different combinations of remotely-sensed environmental variables. Variables included Landsat 8 Band 7, Normalized Difference Vegetation Index (NDVI), land surface temperature (LST) and elevation. Lower scores indicate better predictive ability

Remotely-sensed environmental variables included in model	Log-CPO score
<i>Glossina pallidipes</i>	
Band 7	48.9
NDVI	59.6
LST	57.8
Elevation	59.8
Band 7, elevation	47.8
Band 7, elevation, NDVI	47.1
Band 7, elevation, LST	47.8
Band 7, elevation, NDVI, LST	32.7
<i>Glossina swynnertoni</i>	
Band 7	41.1
NDVI	38.0
LST	42.0
Elevation	41.1
NDVI, elevation	29.3
NDVI, elevation, LST	27.3
NDVI, elevation, Band 7	24.2
NDVI, elevation, Band 7, LST	27.3

**TABLE 2** Posterior estimates of geostatistical model coefficients. Based on fits to  $\log(y + 1)$  transformed data

	M	SD	0.025 quantile	0.5 quantile	0.975 quantile
<i>Glossina pallidipes</i>					
Intercept	12.42	5.180	0.674	12.90	21.34
NDVI	-7.633	4.147	-14.78	-8.023	1.605
Elevation	-0.003	0.002	-0.007	-0.003	0.002
Band 7	-43.10	10.97	-62.87	-43.79	-19.41
LST	0.074	0.096	-0.124	0.076	0.259
Range	3,323	1,814	1,118	2,900	7,985
<i>Glossina swynnertoni</i>					
Intercept	13.09	3.035	6.862	13.15	18.98
NDVI	-10.30	2.723	-15.63	-10.33	-4.793
Elevation	-0.003	0.002	-0.006	-0.004	-0.0004
Band 7	-17.65	6.718	-30.78	-17.71	-4.144
Range	2,477	1,199	951	2,214	5,531

NDVI, Normalized Difference Vegetation Index; Band 7, Landsat 8 Band 7; LST, land surface temperature.

( $\pm 17$ ) in dense woodland (Figure 2a). This was c. 30 times greater than that observed in grassland habitat and six times greater than in savanna habitat (Figure 2a). The mean number of *G. swynnertoni* caught per trap per day was more homogenous across vegetation types (Figure 2b).

For *G. pallidipes*, the model, using remotely-sensed data, with the lowest log-CPO score included all the environmental variables (Table 1). Although the distribution of *G. swynnertoni* appeared homogenous between habitat types (Figure 2b), the model including NDVI, elevation and Band 7 had better predictive ability than other models (Table 1). The mean NDVI between grassland and savanna sites was similar—0.40 ( $\pm 0.01$ ) in grassland and 0.38 ( $\pm 0.01$ ) in savanna which may indicate why similar numbers of *G. swynnertoni* were caught between these two habitats.

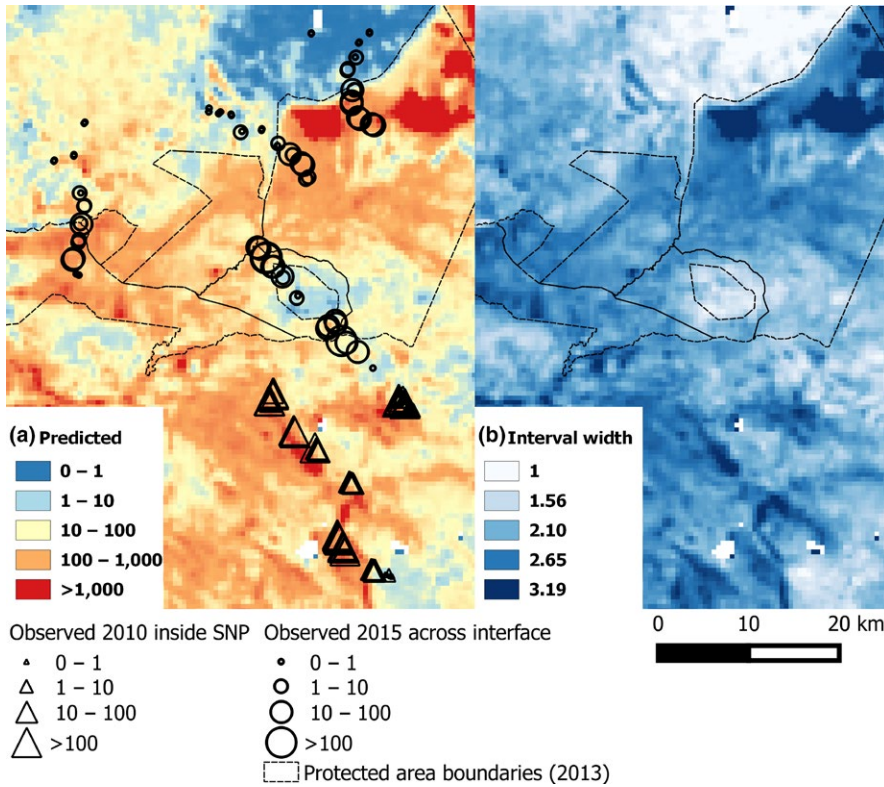
Models including a spatially structured error term provided better fits to the data for both *G. pallidipes* (log-CPO score 25.5) and *G. swynnertoni* (log-CPO score 8.4), than models including an unstructured error term (log-CPO score 37.5 and 23.4). Posterior estimates of model coefficients for the most parsimonious models including a spatially structured error term are provided in Table 2. The distance beyond which abundance was no longer correlated was 3,323 m for *G. pallidipes* and 2,477 m for *G. swynnertoni*. For both species, the amount by which abundance declines as elevation increases was estimated to be equivalent. However, increases in Band 7—indicating lower land surface moisture—was related to a greater decline in *G. pallidipes* abundance than *G. swynnertoni* abundance.

For both tsetse species, predicted abundance for the sites surveyed in 2010 were overestimates for lower observed values and underestimates for higher observed values (Figure S1). This is likely because the relationships between log-transformed abundance and remotely-sensed variables were not exactly linear. However, for both species, maps produced by geostatistical model fits show that most areas predicted to have an abundance >100 were inside the protected area (Figures 3a and 4a). In addition, to the north of Ikorongo Game Reserve, there was a clear change in predicted abundance along the protected area boundary. Inside the protected area, the predicted abundance was >100, but within 5 km into unprotected areas, the predicted abundance was <10. Some unprotected areas to the north of the Grumeti Game Reserve had predicted values of >100 (Figure 3a). *Glossina swynnertoni* was also predicted to be present with abundance >100 in the unprotected areas along the SNP transect (Figures 1 and 4a).

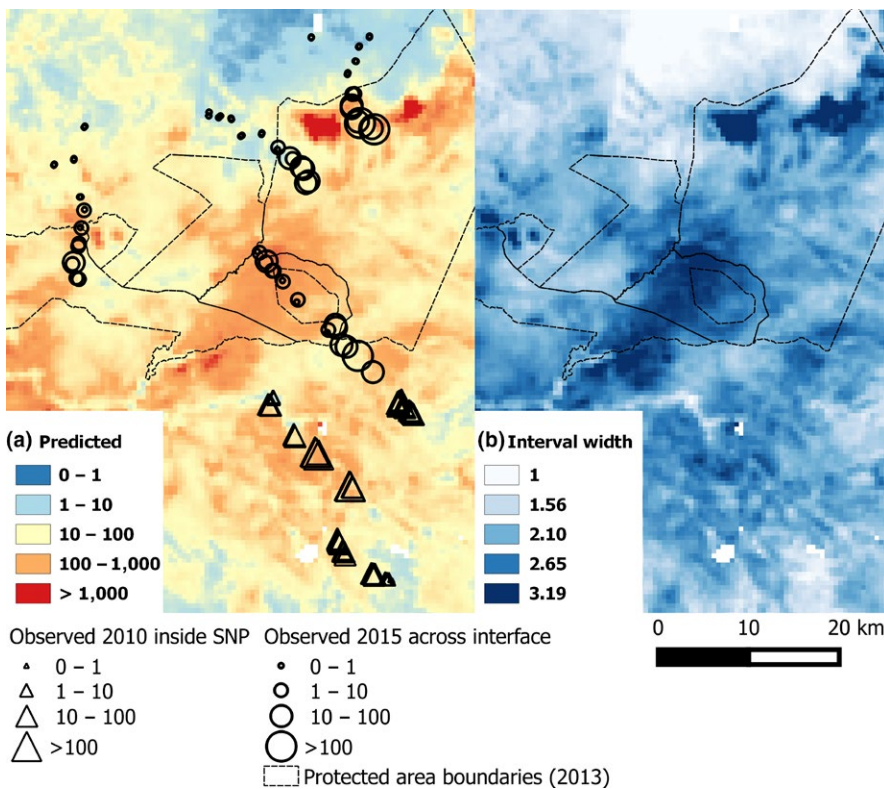
#### 3.2 | Tsetse abundance at the interface between protected and unprotected areas

Mean values of the environmental variables were similar between 2010 trap sites inside the SNP and the 2015 trap sites across the interface (Table S1). Therefore, the two areas were assumed comparable with respect to the remotely-sensed variables. At the interface during 2015, on all transects, the abundance of *G. swynnertoni* and *G. pallidipes* declined to zero by 5 km outside the boundary of the protected area in February 2015 (Figure 5a,b). In general, the geostatistical models of abundance also predicted a decline from inside to outside protected





**FIGURE 3** Predictive map of *Glossina pallidipes* abundance based on geostatistical models fitted to 2010 data >10 km inside the Serengeti National Park (triangles). (a) Bayesian posterior mean predicted values, circles—abundance observed during 2015 across the protected area boundary and (b) Bayesian credible interval width ( $\log_{10}$ )—larger values indicating greater model uncertainty in predicted values. See Figure 1 for details on locations of protected areas



**FIGURE 4** Predictive map of *Glossina swynnertoni* abundance based on geostatistical models fitted to 2010 data >10 km inside the Serengeti National Park (triangles). (a) Bayesian posterior mean predicted values; and (b) Bayesian credible interval width ( $\log_{10}$ )—higher values indicating greater model uncertainty in predicted values. See Figure 1 for details on locations of protected areas

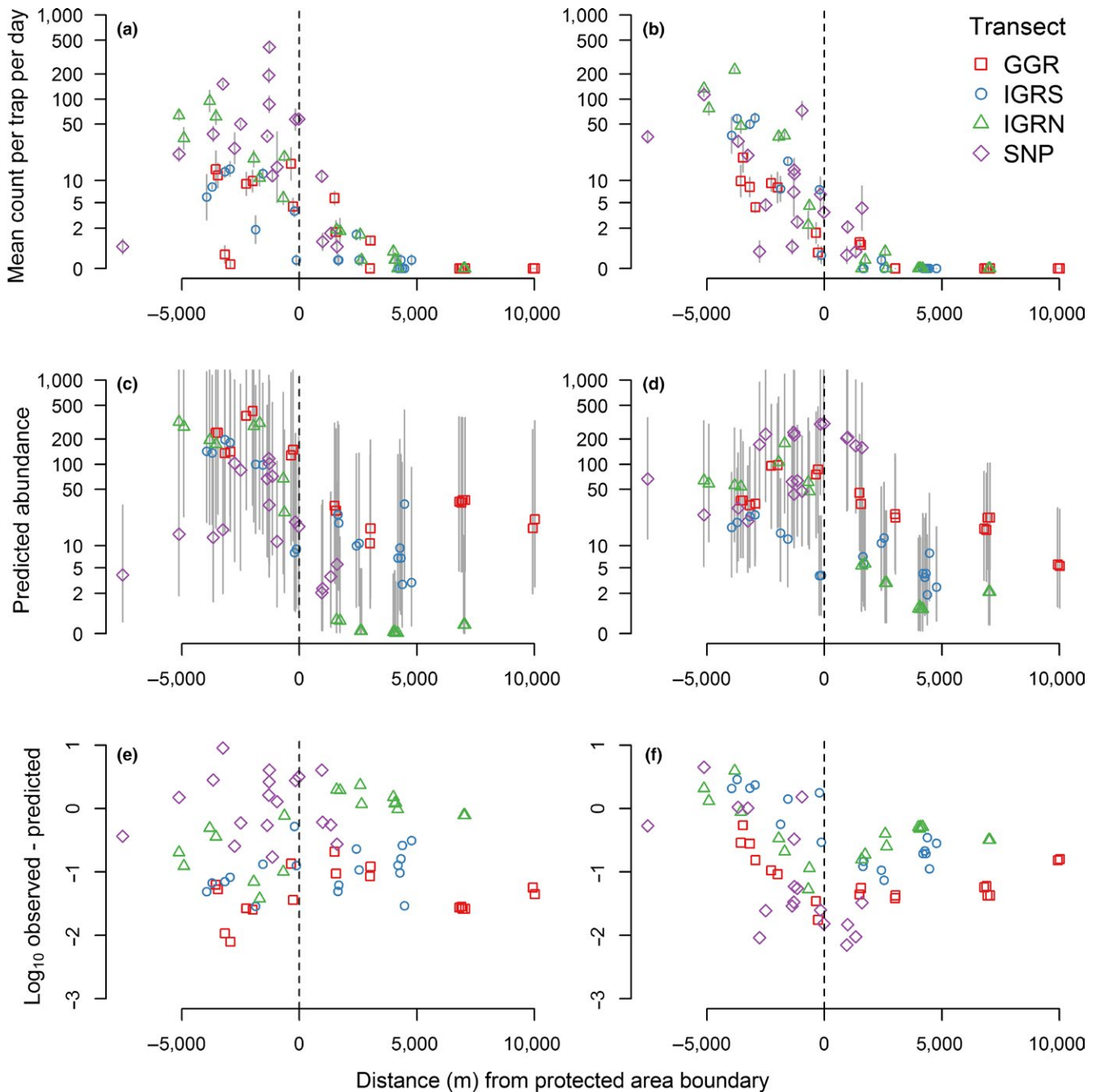
areas (Figure 5c,d). Indeed, the protected area boundary to the north of Ikorongo Game Reserve is quite clearly depicted in the predictive map for *G. pallidipes* and similarly so for *G. swynnertoni* (Figures 3a and 4a), even though the models were only fitted to data >10 km inside the SNP.

Although a general decline was predicted, model predictions of abundance were over estimates for approximately 80% of trap sites (*G. swynnertoni*—59/72, *G. pallidipes*—55/72; Figure 5e,f). The only locations predicted to have abundance >10, more than 5 km into unprotected

areas, were along the Grumeti Game Reserve transect (Figure 5c,d). Field observations confirmed that along this transect there was woody vegetation that may have been sufficient to support tsetse (Figure S2). However, beyond 5 km from the protected area boundary, no tsetse were caught along any of the transects during 2015 surveys. Sites near the boundary where *G. pallidipes* abundance was on average over 50 per trap per day were all associated with riparian vegetation.

## 4 | DISCUSSION

Geostatistical models using remotely-sensed data have potential to identify sites of relatively high vector abundance, for surveillance and control, beyond the spatial extent of initial sampling. We have tested this approach in the case of tsetse flies, by fitting models to the abundance of flies >10 km inside the SNP and testing model predictions



**FIGURE 5** Observed (a, b) and predicted (c, d) tsetse abundance and model residuals (observed–predicted) (e, f) at the interface between protected and unprotected areas. (a) *Glossina pallidipes* observed; (b) *Glossina swynnertoni* observed; (c) *G. pallidipes* predicted; (d) *G. swynnertoni* predicted; (e) *G. pallidipes* residuals; and (f) *G. swynnertoni* residuals. Geostatistical models fitted to 2010 data from >10 km inside the Serengeti National Park. Grey lines in (a) and (b)—SE, grey lines in (c) and (d)—95% credible intervals. Negative distance values on the x-axis indicate locations inside the protected area. GGR, Grumeti Game Reserve; IGRS, Ikorongo Game Reserve South; IGRN, Ikorongo Game Reserve North; SNP, Serengeti National Park. For map of trap locations see Figure 1

across the interface between protected and unprotected areas. In general, the models fitted to data >10 km inside the SNP, predicted a decline in tsetse abundance across the interface. A decline was confirmed by field sampling.

Similar declines in *G. morsitans* have been noted in Zambia and Malawi (Ducheyne et al., 2009; Gondwe et al., 2009; Mweempwa et al., 2015). Our results indicate that loss of woody vegetation explains, in part, such a decline in northern Serengeti, Tanzania. However, our models highlight there may be regions outside protected areas where vegetation is still sufficient to support tsetse populations. Our approach could be used to assist in identifying these remaining regions on a local scale, so that control or monitoring can be targeted effectively.

In our study, sites near the boundary where *G. pallidipes* abundance was on average >50 per trap per day were all associated with riparian woody vegetation. Rivers running from inside to outside protected areas may be where this species could encounter humans and livestock. Control of *G. pallidipes* in this region using insecticide-treated targets should therefore focus on these riparian habitats.

While *G. pallidipes* was found in higher numbers in dense woodland inside the SNP, the abundance of *G. swynnertoni* did not differ significantly between habitat types. This may be due to the actual density of *G. swynnertoni* being similar between habitat types, or it may be due to trap bias. Indeed, odour-baited stationary traps are biased towards host-seeking flies (Hargrove & Packer, 1993). However, the degree of bias is likely different for *G. pallidipes* compared with *G. swynnertoni*. *Glossina pallidipes* is more available to stationary hosts and odour-baited traps than *G. morsitans* and the latter is more attracted to mobile visual baits relative to stationary ones (Vale, 1974). The resting sites of both species are associated with woody vegetation (Chadwick, 1963; Pilson & Leggate, 1962). If the majority of *G. swynnertoni* and *G. pallidipes* rest in woody vegetation, but a smaller proportion of resting *G. swynnertoni* are stimulated by stationary odour-baited traps, then we may overestimate the relative importance of *G. pallidipes*.

Although the abundance of *G. swynnertoni* was more homogenous between vegetation types than *G. pallidipes*, models still predicted a decline in numbers from protected to unprotected areas. Remotely-sensed variables may not correlate directly with habitat type and may reflect other vegetation characteristics important to tsetse. For example, long and short grassland were both grouped as grassland but may have different NDVI values. The similar abundance of *G. swynnertoni* in grassland and savanna habitats may also be due, in part, to the context within which grassland sampling sites were situated with respect to woodland, which we did not account for.

It may be anticipated that regions adjacent to protected areas are subject to relatively high numbers of tsetse dispersing from inside protected areas. However, numbers were already greatly reduced—to 5% of the maximum count in our study space—at the boundary. Our models explained, in part, this observed decline in tsetse abundance, but predicted tsetse to be present in numbers >100 in some unprotected areas. The presence of suitable vegetation at these locations was confirmed by field observations. Even in these regions, abundance was

usually <10 at distances 1 km from the boundary, and catches were zero at >5 km from the boundary.

Models generally overestimated abundance at the interface. Overestimates and the predicted more gradual decline across the interface may be expected as the geostatistical models do not account for context. This may also be due, in part, to factors affecting tsetse population dynamics in different ways between 2010 and 2015. A potential limitation to the study is that there was a lack of contemporary Landsat imagery corresponding to the area surveyed in 2010 situated >10 km inside the SNP. The use of 2015 Landsat data to represent conditions in 2010 may therefore explain the model overpredictions in tsetse abundance observed across the interface in 2015, as seasonal conditions in the preceding months may have differed in 2010 in comparison to 2015. However, as the land within the SNP is protected, we expect these changes to be relatively minimal, and unlikely to have a large impact on tsetse habitat. This is supported by the results of the model both with respect to the relationships observed between 2015 Landsat data and 2010 observed relative abundance and in the validation of the model predictions.

The relationships between vector abundance and remotely-sensed environmental variables are indirect and likely complex, which are not reflected in our linear models. This likely explains why model uncertainty increased with model predicted values. However, the goal of the study was not to predict exact numbers of tsetse, but to find general patterns and establish relative differences between areas that can help to guide surveillance efforts.

Vector-borne disease control programmes often include reducing the life span of the vectors (Rozendaal, 1997; WHO Expert Committee on Control and Surveillance of Human African Trypanosomiasis, 2013; World Health Organization Global Malaria Programme, 2007). Local spatial variation in vector population dynamics influences the efficiency and success of that strategy (Lambrechts, Knox, Wong, Liebman, &, 2009; Ostfeld, Glass, & Keasing, 2005). Knowledge of local variation in vector abundance is therefore essential in planning the control operation, particularly when a disease is close to elimination or incidence is focal. Vector surveillance at scales relevant to control programmes, but sufficient to quantify local-scale variation, is however difficult due to logistics, limited resources and costs of intensive sampling (Alimi et al., 2015). The ability to correlate vector abundance with remotely-sensed data can help direct limited resources (Kalluri et al., 2007). However, the majority use spatial interpolation—predictions made within the same area as the measured data—rather than extrapolation. Extrapolation and validation of predictions within surveillance efforts would require less intensive sampling than grid-based sampling spanning an entire area planned for control.

We used contemporary open-source remote-sensing data. Our approach did not require extensive ground-based information. The method may therefore be suitable for providing initial local-scale predictions over 100s km<sup>2</sup>. In addition, given rapid land-use change occurring at the edges of protected areas (Wittemyer et al., 2008), using models based on remotely-sensed data which is updated monthly is advantageous.



The resulting models will help managers of vector control programmes to select sites, predicted to have relatively high vector abundance, for surveillance prior to targeting control efforts. This will reduce the chances that suitable sites are missed. Models should then be tested by selecting areas predicted to support relatively high vs. low abundance and these sites subject to tsetse surveillance. This will help guide surveillance to targeted areas, and lead to more efficient intervention efforts.

## ACKNOWLEDGEMENTS

The authors thank the field staff at Vector and Vector-Borne Diseases Research Institute, Tanzania. Permits were acquired from COSTECH and TAWIRI, Tanzania. This work was funded by the Biotechnology and Biological Sciences Research Council, the Department for International Development, The Economic and Social Science Research Council, The Natural Environment Research Council and the Defence, Science and Technology Laboratory, under the Zoonosis and Emerging and Livestock Systems (ZELS) programme (grant no. BB/L019035/1) and the UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), the Canadian International Development Research Centre (IDRC), and by the European Union's Seventh Framework Program (FP7/2007-2013) under grant agreement 221948, ICONZ (Integrated Control of Neglected Zoonoses). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## AUTHORS' CONTRIBUTIONS

J.S.L., S.J.T., H.K.A. and L.J.M. jointly conceived the study. J.S.L., S.J.T., H.K.A., M.B. and F.M. were involved in field study design and data collection. J.S.L., S.J.T., M.C.S. and P.M.B. carried out data analysis. All authors contributed to writing and reviews of the manuscript. All authors gave final approval for publication.

## DATA ACCESSIBILITY

Data are available via Zenodo <https://doi.org/10.5281/zenodo.1133429> (Lord et al., 2018).

## ORCID

Jennifer S. Lord  <http://orcid.org/0000-0001-6616-1526>

## REFERENCES

- Albert, M., Wardrop, N. A., Atkinson, P. M., Torr, S. J., & Welburn, S. C. (2015). Tsetse fly (*G. f. fuscipes*) distribution in the Lake Victoria basin of Uganda. *PLoS Neglected Tropical Diseases*, 9, 1–14.
- Alimi, T. O., Fuller, D. O., Quinones, M. L., Xue, R.-D. D., Herrera, S. V., Arevalo-Herrera, M., ... Beier, J. C. (2015). Prospects and recommendations for risk mapping to improve strategies for effective malaria vector control interventions in Latin America. *Malaria Journal*, 14, 519. <https://doi.org/10.1186/s12936-015-1052-1>
- Allsopp, R., Baldry, D. A., & Rodrigues, C. (1972). The influence of game animals on the distribution and feeding habits of *Glossina pallidipes* in the Lambwe Valley. *Bulletin of the World Health Organization*, 47, 795–809.
- Auty, H. (2009). *Ecology of a vector-borne zoonosis in a complex ecosystem: Trypanosomiasis in Serengeti*. Tanzania: University of Edinburgh.
- Auty, H., Anderson, N. E., Picozzi, K., Lembo, T., Mubanga, J., Hoare, R., ... Welburn, S. C. (2012). Trypanosome diversity in wildlife species from the Serengeti and Luangwa valley ecosystems ed J.M. Ndung'u. *PLoS Neglected Tropical Diseases*, 6, e1828. <https://doi.org/10.1371/journal.pntd.0001828>
- Auty, H., Morrison, L. J., Torr, S. J., & Lord, J. (2016). Transmission dynamics of Rhodesian sleeping sickness at the interface of wildlife and livestock areas. *Trends in Parasitology*, 32, 608–621. <https://doi.org/10.1016/j.pt.2016.05.003>
- Barsi, J. A., Lee, K., Kvaran, G., Markham, B. L., & Pedelty, J. A. (2014). The spectral response of the Landsat-8 operational land imager. *Remote Sensing*, 6, 10232–10251. <https://doi.org/10.3390/rs61010232>
- Bouyer, J., Seck, M. T., Sall, B., Ndiaye, E. Y., Guerrini, L., & Vreysen, M. J. B. (2010). Stratified entomological sampling in preparation for an area-wide integrated pest management program: The example of *Glossina palpalis gambiensis* (Diptera: Glossinidae) in the Niayes of Senegal. *Journal of Entomology*, 47, 543–552.
- Brown, P. E. (2015). Model-based geostatistics the easy way. *Journal of Statistical Software*, 63, 1–24.
- Cecchi, G., Paone, M., Feldmann, U., Vreysen, M. J. B., Diall, O., & Mattioli, R. C. (2014). Assembling a geospatial database of tsetse-transmitted animal trypanosomosis for Africa. *Parasites & Vectors*, 7, 39. <https://doi.org/10.1186/1756-3305-7-39>
- Chadwick, P. R. (1963). A study of the resting sites of *Glossina swynnertoni* Aust. in Northern Tanzania. *Bulletin of Entomological Research*, 55, 23–28.
- Davey, B. Y. J. B. (1924). The outbreak of human trypanosomiasis (*Trypanosoma rhodesiense* infection) in Mwanza district, Tanganyika territory. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 17, 474–481. [https://doi.org/10.1016/S0035-9203\(23\)91181-8](https://doi.org/10.1016/S0035-9203(23)91181-8)
- Diall, O., Cecchi, G., Wanda, G., Argilés-Herrero, R., Vreysen, M. J. B., Cattoli, G., ... Bouyer, J. (2017). Developing a progressive control pathway for African animal trypanosomosis. *Trends in Parasitology*, 33, 499–509. <https://doi.org/10.1016/j.pt.2017.02.005>
- Dicko, A. H., Lancelot, R., Seck, M. T., Guerrini, L., Sall, B., Lo, M., ... Bouyer, J. (2014). Using species distribution models to optimize vector control in the framework of the tsetse eradication campaign in Senegal. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 10149–10154. <https://doi.org/10.1073/pnas.1407773111>
- Diggle, P. J., Tawn, J. A., & Moyeed, R. A. (1998). Model-based geostatistics. *Applied Statistics*, 47, 299–350.
- Ducheyne, E., Mweempwa, C., De Pus, C., Vernieuwe, H., De Deken, R., Hendrickx, G., & Van den Bossche, P. (2009). The impact of habitat fragmentation on tsetse abundance on the plateau of eastern Zambia. *Preventive Veterinary Medicine*, 91, 11–18. <https://doi.org/10.1016/j.prevetmed.2009.05.009>
- Estes, A. B., Kuemmerle, T., Kushnir, H., Radeloff, V. C., & Shugart, H. H. (2012). Land-cover change and human population trends in the greater Serengeti ecosystem from 1984–2003. *Biological Conservation*, 147, 255–263. <https://doi.org/10.1016/j.biocon.2012.01.010>
- Glasgow, J., & Welch, J. (1962). Long-term fluctuations in numbers of the tsetse fly *Glossina swynnertoni* Austen. *Bulletin of Entomological Research*, 53, 129–137. <https://doi.org/10.1017/S000748530004801X>
- Gondwe, N., Marcotty, T., Vanwambeke, S. O., De Pus, C., Mulumba, M., & Van Den Bossche, P. (2009). Distribution and density of tsetse flies (*Glossinidae*: Diptera) at the game/people/livestock interface of the Nkhotakota Game Reserve human sleeping sickness focus in Malawi. *EcoHealth*, 6, 260–265. <https://doi.org/10.1007/s10393-009-0252-y>

- Hargrove, J. (1981). Tsetse dispersal reconsidered. *Journal of Animal Ecology*, 64, 351–373. <https://doi.org/10.2307/4060>
- Hargrove, J. W. (1988). Tsetse: The limits to population growth. *Medical and Veterinary Entomology*, 2, 203–217. <https://doi.org/10.1111/j.1365-2915.1988.tb00184.x>
- Hargrove, J. W. (2004). Tsetse population dynamics. In I. Maudlin, P. Holmes & M. Miles (Eds.), *The Trypanosomiasis* (pp. 113–135). Oxford: CABI Publishing. <https://doi.org/10.1079/9780851994758.0000>
- Hargrove, J. W., & Packer, M. J. (1993). Nutritional states of male tsetse flies (*Glossina* spp.) (Diptera: Glossinidae) caught in odour-baited traps and artificial refuges: Models for feeding and digestion. *Bulletin of Entomological Research*, 83, 29. <https://doi.org/10.1017/S0007485300041754>
- Hassell, J. M., Begon, M., Ward, M. J., & Fèvre, E. M. (2016). Urbanization and disease emergence: Dynamics at the wildlife–livestock–human interface. *Trends in Ecology & Evolution*, 32, 55–67. <https://doi.org/10.1016/j.tree.2016.09.012>
- Held, L., Schrödle, B., & Rue, H. (2010). Posterior and cross-validated predictive checks: A comparison of MCMC and INLA. In T. Kneib & G. Tutz (Eds.), *Statistical modelling and regression structures* (pp. 91–110). Heidelberg: Physica Verlag. <https://doi.org/10.1007/978-3-7908-2413-1>
- Hendrickx, G., Napala, A., Dao, B., Batawui, D., De Deken, R., Vermeilen, A., & Slingenbergh, J. (1999). A systematic approach to area-wide tsetse distribution and abundance maps. *Bulletin of Entomological Research*, 89, 231–244.
- Hopcraft, G. (2008). Greater serengeti ecosystem 1:250 000 scale map. Frankfurt Zoological Society, Harvey Map Services Ltd.
- Kaare, M. T., Picozzi, K., Mlengeya, T., Fèvre, E. M., Mellau, L. S., Mtambo, M. M., ... Welburn, S. C. (2007). Sleeping sickness – A re-emerging disease in the Serengeti? *Travel Medicine and Infectious Disease*, 5, 117–124. <https://doi.org/10.1016/j.tmaid.2006.01.014>
- Kalluri, S., Gilruth, P., Rogers, D., & Szczur, M. (2007). Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: A review. *PLoS Pathogens*, 3, 1361–1371.
- Kitron, U., Otieno, L. H., Hungerford, L. L., Odulaja, A., Brigham, W. U., Okello, O. O., ... Cook, E. (1996). Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS. *The Journal of Animal Ecology*, 65, 371. <https://doi.org/10.2307/5883>
- Lambrechts, L., Knox, T. B., Wong, J., Liebman, K. A., Albright, R. G., & Stoddard, S. T. (2009). Shifting priorities in vector biology to improve control of vector-borne disease. *Tropical Medicine & International Health*, 14, 1505–1514. <https://doi.org/10.1111/j.1365-3156.2009.02401.x>
- Lin, S., DeVisser, M. H., & Messina, J. P. (2015). An agent-based model to simulate tsetse fly distribution and control techniques: A case study in Nguruman, Kenya. *Ecological Modelling*, 314, 80–89. <https://doi.org/10.1016/j.ecolmodel.2015.07.015>
- Lord, J., Torr, S., Auty, H., Brock, P., Byamungu, M., Hargrove, J., ... Stanton, M. (2018). Data from: Geostatistical models using remotely-sensed data predict savanna tsetse decline across the interface between protected and unprotected areas in Serengeti, Tanzania. *Zenodo Digital Repository*, <https://doi.org/10.5281/zenodo.1133429>
- Mihok, S. (2002). The development of a multipurpose trap (the Nzi) for tsetse and other biting flies. *Bulletin of Entomological Research*, 92, 385–403.
- Miller, J., Turner, M., Smithwick, E., Dent, C., & Stanley, E. (2004). Spatial extrapolation: The science of predicting ecological patterns and processes. *BioScience*, 54, 310. [https://doi.org/10.1641/0006-3568\(2004\)054\[0310:SETSOP\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0310:SETSOP]2.0.CO;2)
- Moore, N., & Messina, J. (2010). A landscape and climate data logistic model of tsetse distribution in Kenya. *PLoS ONE*, 5, e11809. <https://doi.org/10.1371/journal.pone.0011809>
- Mweempwa, C., Marcotty, T., De Pus, C., Penzhorn, B. L., Dicko, A. H., Bouyer, J., & De Deken, R. (2015). Impact of habitat fragmentation on tsetse populations and trypanosomosis risk in Eastern Zambia. *Parasites & Vectors*, 8, 406. <https://doi.org/10.1186/s13071-015-1018-8>
- Odiit, M., Coleman, P. G., Liu, W. C., McDermott, J. J., Fèvre, E. M., Welburn, S. C., & Woolhouse, M. E. J. (2005). Quantifying the level of under-detection of *Trypanosoma brucei rhodesiense* sleeping sickness cases. *Tropical Medicine & International Health*, 10, 840–849. <https://doi.org/10.1111/j.1365-3156.2005.01470.x>
- Ostfeld, R. S., Glass, G. E., & Keasing, F. (2005). Spatial epidemiology: An emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, 20, 328–336. <https://doi.org/10.1016/j.tree.2005.03.009>
- Pilson, R., & Leggate, B. (1962). A diurnal and seasonal study of the resting behaviour of *Glossina pallidipes* Aust. *Australian Bulletin of Entomological Research*, 53, 551–562. <https://doi.org/10.1017/S0007485300048318>
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: The R Foundation for Statistical Computing.
- Reed, D. N., Anderson, T. M., Dempewolf, J., Metzger, K., & Serneels, S. (2009). The spatial distribution of vegetation types in the Serengeti ecosystem: The influence of rainfall and topographic relief on vegetation patch characteristics. *Journal of Biogeography*, 36, 770–782. <https://doi.org/10.1111/j.1365-2699.2008.02017.x>
- Rogers, D., & Randolph, S. (1991). Mortality rates and population density of tsetse flies correlated with satellite imagery. *Nature*, 351, 739–741. <https://doi.org/10.1038/351739a0>
- Rozendaal, J. (1997). *Vector control: Methods for use by individuals and communities*. Geneva, Switzerland: World Health Organization.
- Rue, H., Martino, S., & Nicolas, C. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, 71, 319–392. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>
- Shaw, A. P. M., Wint, G. R. W., Cecchi, G., Torr, S. J., Mattioli, R. C., & Robinson, T. P. (2015). Mapping the benefit-cost ratios of interventions against bovine trypanosomosis in Eastern Africa. *Preventive Veterinary Medicine*, 122, 406–416. <https://doi.org/10.1016/j.prevetmed.2015.06.013>
- Simarro, P. P., Cecchi, G., Paone, M., Franco, J. R., Diarra, A., Ruiz, J. A., ... Jannin, J. G. (2010). The Atlas of human African trypanosomiasis: A contribution to global mapping of neglected tropical diseases. *International Journal of Health Geographics*, 9, 57. <https://doi.org/10.1186/1476-072X-9-57>
- Sobrinho, J. A., Jiménez-Muñoz, J. C., & Paolini, L. (2004). Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment*, 90, 434–440. <https://doi.org/10.1016/j.rse.2004.02.003>
- Swynnerton, C. (1923). The entomological aspects of an outbreak of sleeping sickness near Mwanza, Tanganyika territory. *Bulletin of Entomological Research*, 13, 317–370.
- Torr, S. J., Hall, D. R., Phelps, R. J., & Vale, G. A. (1997). Methods for dispensing odour attractants for tsetse flies (Diptera: Glossinidae). *Bulletin of Entomological Research*, 87, 299. <https://doi.org/10.1017/S0007485300037251>
- Vale, G. A. (1974). The responses of tsetse flies (Diptera, Glossinidae) to mobile and stationary baits. *Bulletin of Entomological Research*, 64, 545. <https://doi.org/10.1017/S0007485300035860>
- Vale, G. A., Hargrove, J. W., Cockbill, G. F., & Phelps, R. J. (1986). Field trials of baits to control populations of *Glossina morsitans morsitans* Westwood and *G. pallidipes* Austen (Diptera: Glossinidae). *Bulletin of Entomological Research*, 76, 179. <https://doi.org/10.1017/S000748530001467X>
- Wint, W., & Rogers, R. (2000). Predicted distributions of Tsetse in Africa. Consultants Report. Prepared by ERGO Ltd and the TALA Research Group Oxford for the Food and Agriculture Organization of the United Nations.
- Wittemyer, G., Elsen, P., Bean, W. T., Coleman, A., Burton, O., & Brashares, J. S. (2008). Accelerated human population growth at protected area edges. *Science*, 321, 123–126. <https://doi.org/10.1126/science.1158900>

- World Health Organization Global Malaria Programme. (2007). *Malaria elimination: A field manual for low and moderate endemic countries*. Geneva, Switzerland: World Health Organization.
- WHO Expert Committee on Control and Surveillance of Human African Trypanosomiasis. (2013). *Control and surveillance of human African trypanosomiasis*. World Health Organization Technical Report Series, 1–237. Geneva, Switzerland: World Health Organization.

**How to cite this article:** Lord JS, Torr SJ, Auty HK, et al. Geostatistical models using remotely-sensed data predict savanna tsetse decline across the interface between protected and unprotected areas in Serengeti, Tanzania. *J Appl Ecol.* 2018;00:1–11. <https://doi.org/10.1111/1365-2664.13091>

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.